



Project Acronym: **GAIN**

Project Full Title: **Georgian Artificial Intelligence Networking and Twinning Initiative**

Grant Agreement Number: 101078950

HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning)

WEBSITE: [www.gain-twinning.eu](http://www.gain-twinning.eu)

Project Start Date: 1<sup>st</sup> October 2022

**Deliverable D.6.3. Data Management Plan, Version 1**

Dissemination level:	Sensitive
Date:	December 31, 2022
Deliverable leader:	MICM
Author:	George Giorgobiani
Contributors:	All
Reviewers:	J. Alexandersson, A. Konig, F. Bremond, P. Müller
Type:	Report
WP / Task responsible:	MICM
Key words:	FAIR, GDPR, ELSI, data, management, plan

## Contents

1. Executive Summary .....	4
2. Introduction .....	4
3. Versions and Updates .....	5
4. Open Dissemination Strategy.....	5
5. Protection of Personal data .....	6
6. FAIR Data Strategy .....	6
<b>6.1. Findability</b> .....	7
<b>6.2. Accessibility</b> .....	9
6.3. Interoperability .....	9
6.4. Reusability .....	10
7. Restricted Data .....	10
8. Zenodo Platform .....	11
References .....	13

## 1. Executive Summary

This document describes the **Data Management Plan, Version 1 (D.6.3.)** of the project GAIN.

Document History			
Version	Date	Contributors	Description
1	27/02/2023	J. Alexandersson, P. Müller, A. Girenko – DFKI; F. Bremond, A. König – INRIA; G. Giorgobiani, Ts. Javakhishvili - MICM	Data Management Plan, Version 1
2	12/03/2023	J. Alexandersson, P. Müller, A. Girenko – DFKI; F. Bremond, A. König – INRIA; G. Giorgobiani, Ts. Javakhishvili - MICM	Data Management Plan, Version 1
3 (final)	25/03/2023	J. Alexandersson, P. Müller, A. Girenko – DFKI; F. Bremond, A. König – INRIA; G. Giorgobiani, Ts. Javakhishvili - MICM	Data Management Plan, Version 1

## 2. Introduction

Data Management Plan, Version 1 is the part of the WP6, Coordination and Management. It gives an overview on the data generation and collection process, how it is documented and saved within the GAIN project. The Plan is based on the 4 principles of FAIR Data (see e.g. [1, 2] in References):

1. Findability
2. Accessibility
3. Interoperability
4. Reusability.

GAIN addresses open science and research data principles as a keystone in advancing EU research. The project will produce:

- Scientific results that will be published in international journals and conferences
- Scientific data generated within the project and published in scientific articles
- Other types of data, including business data collected from the partners, such as deliverables, meeting minutes, technical documents and other.

For sharing the results “Project’s Online Management System”, the project website and the Zenodo Platform<sup>1</sup> will be used. The latter fully complies with the FAIR principles (see section 8 below).

Important project (baseline) elements, which will be used for joint research and GAIN capacity building, include the creation of a multimodal corpora (speech, video, biosensors) of **social patient-clinician interactions**. Technical tools and organizational methods for the management of this kind of data will include the aspects of ethical, legal and social implications (**ELSI**), **GDPR** requirements compliance, Consortium Agreement of GAIN and the **Ethics Protocol** (Deliverable 1.2) created within the project.

### 3. Versions and Updates

According to the DoW of the project, the data management plan will be updated at the middle of the project as version 2 (Deliverable 6.4), which, at the same time, will be a final version of the deliverable. The reasons of updates might be the significant changes in the project, such as:

- Availability of new data sources
- Changes in the project consortium
- Other reasons that might be of relevance for the project.

The data management plan will be reviewed and updated as given in Table 1:

Version	Content	Due date
1	Version 1 of the data management plan will take into account the requirements of the General Data Protection Regulation	Month 6
2	Version 2 will be the update of the data management plan at the middle of the project and a final version as well	Month 18

Table 1

### 4. Open Dissemination Strategy

The project dissemination and communication activities are designed to maximally contribute to attaining the overall project goal – to integrate MICM into the community of leading European research organizations in the field of AI. Therefore, the project dissemination will be focused on spreading the

---

<sup>1</sup> <https://zenodo.org/>

research results obtained through pursuing the TRIP<sup>2</sup> and address the research audience. GAIN consortium has reserved funds for covering the publication fees so as to ensure that open access will be provided to the publications (papers, the book). GAIN will endeavour gold open access and will also ensure that the articles reporting results will be published in journals with high impact factor and coverage.

In cases, when the gold model is not applicable (such if the budget for open access is exhausted), the open access will be given according to the “Green Model” (publications are made available on the project website) after a possible embargo period.

## 5. Protection of Personal data

Only the personal data that might be necessary for the project will be collected. The data collected in the GAIN project will be securely stored. Personal data protection will follow the Directive 95/46/EC of the GDPR.

Personal identifiers will be protected through pseudonymization. According to the Article 4, paragraph 1 of the GDPR, “personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.”

Transmission of the data to entities outside of the EU will be regulated by the Chapter 5, Transfers of personal data to third countries or international organisations, of GDPR and the project’s Consortium Agreement.

## 6. FAIR Data Strategy

This part concerns the project data such as scientific results that will be published in international journals and conferences, scientific data generated within the project and published in scientific articles, other types of data, including business data collected from the partners, such as deliverables, meeting minutes, technical documents and other.

As pointed out in the introduction, Data Management Plan, Version 1 is based on FAIR data principles as given in [1] and provides guidelines for the Findability, Accessibility, Interoperability and Reuse of the project generated data.

---

<sup>2</sup> TRIP - Twinning Research and Innovation Programme

## 6.1. Findability

The FAIR Guiding Principles for the Findability are

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource.

Data produced in the GAIN project will be shared mainly through the **Project Online Management Platform** (Deliverable D.6.2) developed as the fast and effective communication, coordination and management system of the project. This is the internal project information platform, a documents repository and a ticketing system for the task management, which is structured in the google drive. The main folders contained in the platform are Deliverables, Development, Meetings, Organizational, Publications, Resources. Access to the platform is provided to all participants of the project and may be shared to external stakeholders as well. It doesn't contain Restricted or Sensitive data, like deliverables and medical data, which are not publicly open. Public data will be shared on the project website, social media, like Facebook and LinkedIn, and other dissemination activities. The public data will not contain confidential personal information.

The front page of the documents will contain the Flag of EU, the logos of Horizon Europe and the partner organisations and brief information about the project as below:

Project Acronym: **GAIN**

Project Full Title: **Georgian Artificial Intelligence Networking and Twinning Initiative**

Grant Agreement Number: 101078950

HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning)

WEBSITE: [www.gain-twinning.eu](http://www.gain-twinning.eu)

Project Start Date: 1<sup>st</sup> October 2022

Furthermore, the documents will contain the basic metadata for the efficient use of the information related with the document by the project partners. The documents will contain information about the dissemination level, date, deliverable leader, author, contributor and reviewer organisations, document type, WP / Task responsible and keywords. As an example see Table 2 below.

Dissemination level:	PU
Date:	December 31, 2022
Deliverable leader:	MICM
Author:	George Giorgobiani
Contributors:	DFKI, INRIA
Reviewers:	DFKI, INRIA
Type:	Report
WP / Task responsible:	MICM
Keywords	Online Management Platform, Data,...

Table 2

There are 3 possible types of Dissemination level of the document (with the corresponding abbreviations):

- Public (PU)
- Restricted (RE)
- Sensitive (SEN).

For the Type of the document there are 4 options (with the corresponding abbreviations):

- Report (R)
- Software (SW)
- Other (O).

Regarding the document versions, all templates generated in the GAIN projects will contain the history of changes of the document: the current version of the document, the date of change, the name of the contributor as well as a short description. As an example see Table 3 below

Document History			
Version	Date	Contributors	Description
1	25/11/2022	George Giorgobiani - MICM	Define and agree the framework parameters of the TRIP



3 (Final)	30/12/2022	George Giorgobiani – MICM; Alexandra Konig – INRIA; Jan Alexandersson - DFKI	Define and agree the framework parameters of the TRIP
-----------	------------	--	---

Table 3

## 6.2. Accessibility

The FAIR Guiding Principles for the Accessibility are:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

Project data and deliverables that are declared as public will be anonymised and made available to the public and all third parties that request access to them. The anonymised data sets may be used as part of the dissemination activities of the project.

Scientific results that will be published in international journals and conferences, scientific data generated within the project and published in scientific articles (e.g. the field data collected through TRIP – e.g. Georgian interactional data corpus, evaluation data, linguistic models, etc.) will be handled in accordance with the FAIR principles. GAIN intends to provide open access to its scientific data to create multiplicative effect in EU research. This type of data will be accessible openly at the Zenodo Platform and project website as well.

Project will produce other types of data, including business data collected from the partners. Again, open access is the primary intention of GAIN. This will be modified only when and if the responsible partner proves that keeping them private is mandatory for the exploitation of the project results or if the information distribution contradicts respective regulations (e.g. personal information, etc.).

## 6.3. Interoperability

The FAIR Guiding Principles for the Interoperability are:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

Interoperability will be ensured by the project partners by storing all data in an appropriate format that will make data accessible to all professionals who are interested in exploiting the data generated in the project.

The language used in the project will be the one in common use within research and medical sectors, addressing the relevant target audience. Vocabulary will not constitute a barrier for data interoperability and reuse. Authors will only use words that are commonly understood by the sector of the target audience.

#### 6.4. Reusability

The FAIR Guiding Principles for the data to be Reusable are:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards.

Provided anonymized data sets will enable the third parties to use the data generated by the project.

After the end of the project, relevant data that might be necessary for further research will be available at the project website. The consortium will maintain this project data in a reusable way for as long as possible after the end of the project. The duration is depending on the agreements among the project partners.

The project coordinator will be responsible for the quality of the data by assuring that datasets follow the FAIR principles included in this document.

### 7. Restricted Data

As pointed out in the Introduction, not all data will be made openly available. This are the data which will be generated in the project and used for joint research and GAIN capacity building, including the creation of a multimodal corpora (speech, video, biosensors) of social patient-clinician interactions etc. The access to these documents needs to be restricted.

Technical tools and organizational methods for the management of this kind of data will include the aspects of ethical, legal and social implications (ELSI), GDPR requirements compliance, Consortium Agreement of GAIN and the Ethics Protocol (Deliverable 1.2) created within the project.

The project will create the necessary tools to handle data with multiple levels of restrictions. This allows for the creation of users with different access levels, access for specific time frames, access to specific data sets as well as access through secure authentication.

### **Protecting the Data**

We will need two types of protection to keep the sensitive data safe. During the sending and receiving and when it will be held on the server.

Sending/Receiving:

- We will use a reliable virtual private network (VPN) and use only this network for sending and receiving sensitive data
- We'll also add an end-to-end encryption, which is a security method that keeps our communications safe (including speech, videos and photos).

Keeping data safe:

- We will create a dedicated user account and allow rw-access to the files/folders only to this user. (Other users will only be able to access this data only via Pytorch)
- We'll have a well set up Access Logging (Which will be monitored routinely)
- Strong passwords are a must have here
- We'll set up a very safe facility for our server, so that no one can enter the room, except for the aforementioned user and a selected few
- Also threatening with legal actions for good measure.

## **8. Zenodo Platform**

Zenodo (<https://zenodo.org>) is an open repository for all scholarship, enabling researchers from all disciplines to share and preserve their research outputs, regardless of size or format. Free to upload and free to access, Zenodo makes scientific outputs of all kinds citable, shareable and discoverable for the long term [3]. Zenodo is good for:

- Sharing and linking research outputs to datasets and funding
- Citable and discoverable: All uploads get a Digital Object Identifier (DOI); All open content is harvestable via OAI-PMH by third parties
- Supports versioning via a top-level DOI
- Trusted, reliable, safe: Data is stored at CERN kept in multiple online and offline copies

- Includes funding information and makes reporting easier: Zenodo allows you to link uploads to various grants, e.g. of European Commission, National Science Foundation, etc.
- Includes article level metrics: Tracks attention for scholarship uploaded
- Flexible licensing
- Reviewing: Research materials can set to share with reviewers only, and also embargoed.
- Software preservation made simple: Zenodo is integrated with GitHub.
- Supports FAIR principles.

The project GAIN has created an account and the community by the name “Georgian Artificial Intelligence Networking and Twinning Initiative” at Zenodo platform (see Table 4).

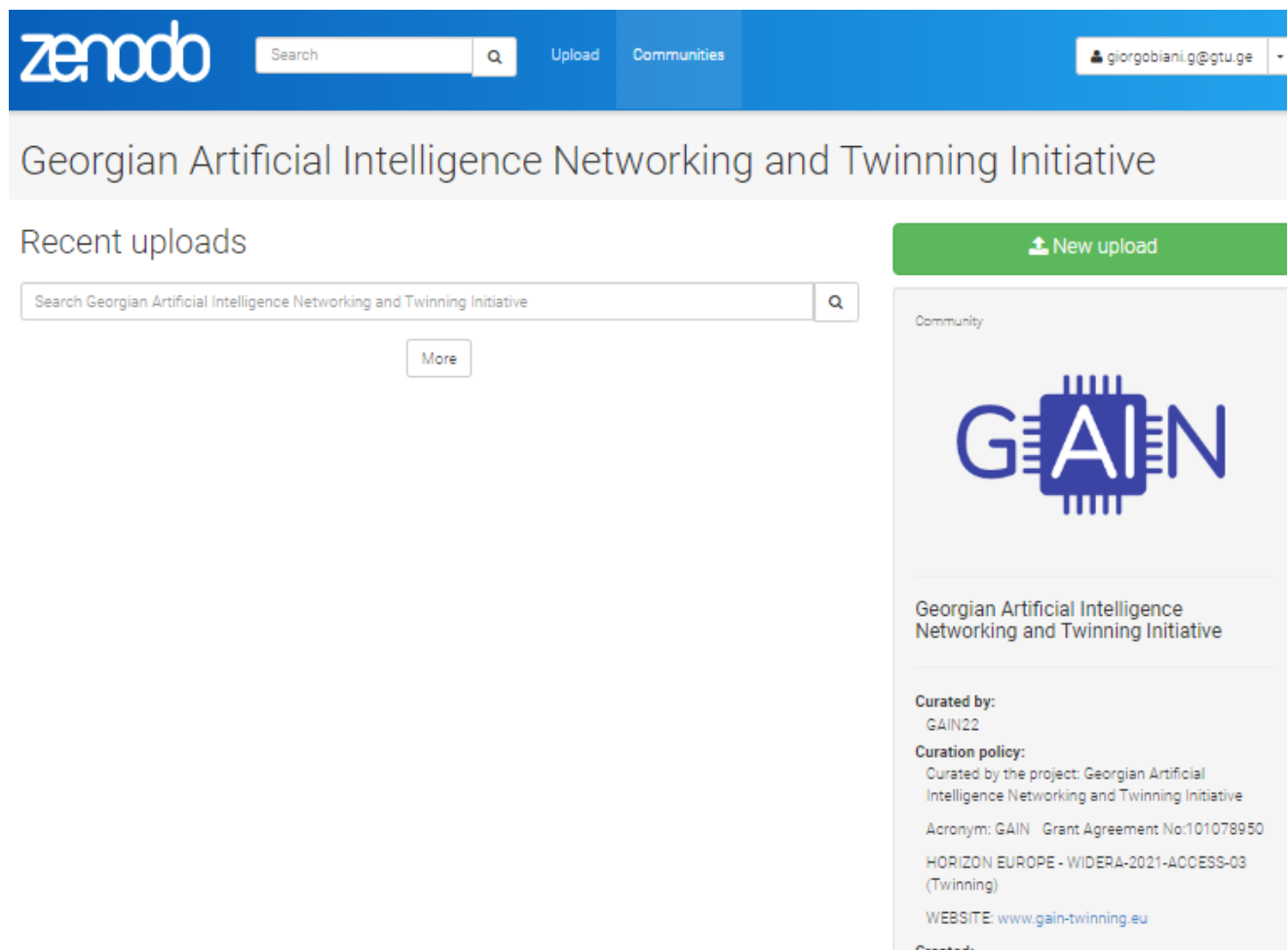


Table 4

## References

1. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
2. <https://www.go-fair.org/fair-principles/>
3. <https://www.openaire.eu/zenodo-guide>